

# Eprint website for the University of Tasmania

Professor Arthur Sale

2004 June 27, Draft Version 0.5

## Executive Summary

This document proposes the urgent establishment of an eprint website for the University of Tasmania, and corresponding policies. A prototype website temporarily codenamed 'UTasER' can be viewed at <http://eprints.comp.utas.edu.au/> from within the University intranet.

## What are eprints?

An *eprint* is an electronic version of a paper, article or thesis, preserved in an archive and searchable and retrievable globally. The word encompasses *preprints* (versions of a paper distributed before refereed publication) and *postprints* or reprints (copies of a published paper distributed by themselves). An eprint server is a server on which all or most of the research output of an institution is mounted, and which provides search and browse capability to find particular papers. Such a server is a useful addition to a university's profile, but not particularly valuable by itself. You have to know about it to search it.

To be really value-adding, an eprint server must comply with the standards of the *Open Access Initiative* (OAI), and be registered with global OAI harvesters such as **myOAI** (<http://www.myoai.com/>) and **OAIster** (<http://oaister.umd.umich.edu/o/oaister/>). These provide global search services for research publications for all registered institutions, for example on Oaister currently 3.2M records from 301 universities and research organizations.

A prototype eprint server has been established for the University of Tasmania and to experiment with the look-and-feel of an eprint server, view it at <http://eprints.comp.utas.edu.au/>. At the date of writing this prototype contains 8 documents: one journal paper, three conference papers, one newspaper article, and three unpublished reports. One report (this document) is available as HTML and has been updated since its first upload; the others are available as PDF files, and one has an alternate Word format. These documents have been chosen to exercise the server's capabilities and provide some idea of the possibilities.

To get some experience, go to the prototype server and search for the key phrase 'spread spectrum' or search for the name 'Malhotra'. This prototype is not public (accessible only on-campus) and therefore not yet registered with the harvesters. Try also viewing OAIster, and searching all 301 institutions for something or someone interesting to you; if your mind is blank try 'spam filter'. To try myOAI you will have to register (free).

***You cannot reasonably comment on this proposal until you have some experience of what it might offer; to assist you this document itself is uploaded to the prototype server as HTML <http://eprints.comp.utas.edu.au:81/archive/00000011/>. Download the HTML version and you will have a set of live hyperlinks that can take you directly to the places on the Web mentioned above and littered through this document.***

The author is willing to give a brief talk and demonstration of eprints (including the prototype server and OAI browsers) on request to [Arthur.Sale@utas.edu.au](mailto:Arthur.Sale@utas.edu.au).

## Benefits

There are many benefits of an eprint website for the University of Tasmania. Perhaps the most significant to academics, RHD students and other researchers are the following which align firmly with the EDGE agenda.

- Our research output is made publicly available, globally, free, and at the time of creation. It is not restricted to an institution, country or journal, or ability to pay.
- The self-loading of preprints on the server provides *prima facie* proof of priority of the research findings. This is especially important for research higher degree theses and is a win-win situation for postgraduates.
- Global searches through the OAI bring our research and researchers more easily to the attention of other researchers worldwide.
- Papers available online are suggested by some information science researchers to be cited on average 300% more frequently than papers available only in paper form! See 'Articles freely available online are more highly cited', *Nature*, **411**-6837, p521, 2001, also at <http://www.neci.nec.com/~lawrence/papers/online-nature01/>.
- Theses have limited availability to anyone outside the originating institution. Placing all publicly accessible theses of RHD graduates on an eprint server provides global access and establishes research priority. For example MIT has done this, and its research is now much more accessed.

Besides these, there are many more peripheral or long-range benefits that are unlikely to motivate academic staff yet which may resonate with senior management. These include:

- The Group of Eight universities have a project to install open access (eprint) archives in all of their membership. At the time of writing only the servers at Melbourne ([www.unimelb.edu.au](http://www.unimelb.edu.au), 260 records), Queensland (<http://eprint.uq.edu.au/>, 875 records), ANU (<http://eprints.anu.edu.au/> 2000 records) and Monash University (<http://eprint.monash.edu.au/>, 33 records) are operational. The University of Tasmania regards itself as equal with these universities. QUT also has an operational server, as does ALIA and the National Library of Australia.
- No university has access to the entire world's research. The open access initiative is aimed at making access to research output freely available to all, and joining this initiative incidentally assists in combating the serials pricing crisis.
- Some disciplines are already highly electronic in their dissemination practices, such as Physics and Computing. This trend can only be expected to continue, and an eprint archive will assist the University in maintaining a leading edge reputation.
- The initiative is an operation driven by standards, where global interoperability is seen as vital.

## Implementation Barriers

### **Direct Costs**

The direct costs (cash) are minor. The prototype eprints server is mounted on the same server used by the School of Computing for many other purposes. A dedicated server would cost say \$5000, with ample disk space for records for several years and a better response time. However, initially a fully operational server could be mounted on an existing University web server.

The EPrints software proposed to be used (<http://www.eprints.org/>) is completely free under a GNU open source licence, as are updates and all the supporting software (Apache, MySQL, Perl, etc).

Registration with OAI harvesters is also free. Searches performed on harvesters such as myOAI and OAIster are free apart from Internet traffic costs. The software is widely used by universities for this purpose and there is an active support forum.

### ***Indirect costs***

Indirect costs are more significant and can be broken down into technical support costs, server supervision, and upload costs.

#### **Technical support by ICT personnel**

The initial implementation effort for the prototype has been supplied by the School of Computing. The implementation could be easily transported to another server with minimal staff time. There will however need to be some work put into customizing the site to suit the University's visual standards and desired user interface. Other university sites offer examples. This need not be a large task, indeed could be minimal and evolve with the site. Depending on the upload solution adopted, it would be desirable for IT Resources to write a module to interface to the University LDAP server so that all research staff have automatic upload registration on the server with their email username and password; this might require say a week's work. Ongoing technical support should be minor, and mainly concerned with updates and backups.

#### **Server supervision by information specialists**

The server will require supervision by someone with a research or information science speciality. Regular monitoring will be required to approve uploads, and monitor the quality of the service and the status of the server. Depending on the take-up of the facilities, this might be a moderate or a relatively light load. An upper bound estimate of the effort can be made by assuming that the entire research and thesis output of the University is uploaded to the server annually.

### **Uploads**

Creation of content is the province of academic staff and RHD candidates/graduates. However, there is the additional step of submitting the content (preprint files and in some cases postprints) to the server. Three models are possible:

1. In one, the academic uploads the file and enters the bibliographic information. Experience suggests that the work involved may be 5-10 minutes with a small amount of experience of what is required. This is a tiny fraction of the work involved in producing the paper, and would seem negligible. However, in other institutions, it has been seen as a barrier because it simply does not get done. It is hard to get academics to do work without deadlines. The quality of the metadata may also be variable.
2. In a variation on this theme, one person in each school is responsible for the collection and uploading. This might be the person responsible for PES data entry since much of the information is duplicated. Entry would be smoother, quicker and more reliable, at the expense of some extra liaison with the academic and workload for that person. This solution seems appropriate for RHD theses, regardless of the way individual papers are uploaded.
3. The ultimate in centralization would be to have a single institutional person do the uploading, with papers simply emailed to that person. This has the ultimate in consistency, but also requires a significant change to the duties of that person. Seeking additional information not initially supplied by the academic in the email would constitute a significant part of that load. (The option exists to spread the workload around various subject editors.)

## Participation

The implementation of an eprint server is easy; the hard part is getting near 100% participation by researchers and coverage of institutional output. This can be readily seen by the performance of Australian institutions with eprint servers (from 33 documents at Monash University to a respectable 2000 at ANU). For comparison, MIT has 8000 theses and 4000 papers; Duke University's Historical Sheet Music Archive has 17 000 records. To save rewriting what others have already experienced, here is what the eprint FAQ says about this problem:

### “How can an institution facilitate the filling of its Eprint archives?”

1. Install [OAI-compliant Eprint Archives](#) .
2. Adopt [a university-wide policy](#) that all faculty maintain and update a [standardised online curriculum vitae \(CV\)](#) for [annual review](#).
3. Mandate that [the full digital text of all refereed publications](#) should be deposited in [the University Eprint Archives](#) and linked to their entry in the author's online CV. (Make it clear to all faculty how self-archiving is [in the interest of their own research and standing](#) , maximizing the [visibility](#) , [accessibility](#) and [impact](#) of their work.)
4. Offer [trained digital librarian](#) help in showing faculty how to self-archive their papers in their own [university Eprint Archive](#) (it is very easy).
5. Offer [trained digital librarian](#) help in doing ["proxy" self-archiving](#), on behalf of any authors who feel that they are personally unable (too busy or technically incapable) to self-archive for themselves. They need only supply their digital full-texts in word-processor form: the digital archiving assistants can do the rest (usually only a few dozen keystrokes per paper).
6. A policy of mandated self-archiving for all refereed research output, together with a trained proxy self-archiving service, to ensure that lack of time or skill do not become grounds for non-compliance, are the most important ingredients in [a successful self-archiving program](#). The proxy self-archiving will only be needed to set the first wave of self-archiving reliably in motion. The [rewards of self-archiving](#) -- in terms of [visibility](#) , [accessibility](#) and [impact](#) -- will maintain the momentum once the archive has reached critical mass. And even students can do for faculty the few keystrokes needed for each new paper thereafter.)
7. [Digital librarians, collaborating with web system staff](#) , should be involved in ensuring the proper maintenance, backup, mirroring, upgrading, and migration that ensure the perpetual preservation of the university Eprint Archives. Mirroring and migration should be handled in collaboration with counterparts at all other institutions supporting OAI-compliant Eprint Archives.”

## Copyright

Wherever an eprint server is proposed, many respond ‘But I can't do this, because the journal/conference I publish in won't let me.’ This is largely nonsense, and there is an extensive literature on the reactions and the common objections. See the ‘I worry about...’ section at <http://www.eprints.org/self-faq/>.

In brief, the research and the paper belong to the academic and/or the employing institution prior to publication. At the preprint stage, the author (or the institution) is free to do whatever they want with it. Indeed in many disciplines there was a healthy trade in paper preprints of research articles until electronic archives took over – the most significant and first example is Physics, but there are many others in the sciences and technologies. In other disciplines the paper preprint culture never took off, especially in the humanities. Regardless of the prior existence of a preprint culture, there is no copyright barrier to mounting preprints on an institutional server, right up to the point where the article is accepted and the publisher asks for a license to use the copyright.

At this stage, *all* publishers of journals or conference proceedings ask for some form of copyright license or more assignment of copyright. In the majority of cases the exact form of this is more a matter of tradition than legal requirement, and the publishers are quite happy for preprints and postprints to be mounted on a personal website or institutional eprint server (for example *Nature*). Indeed in the computer sciences, some publishers will provide the postprint PDF file as printed in the paper journal or conference proceedings for the author to mount personally (for example the *Journal of Research & Practice in Information Technology*). The number of publishers that insist on sole rights is decreasing (estimated at 40%), and where possible researchers might consider not submitting to them. For an introduction to the literature on this topic see <http://www.eprints.org/self-faq/#publisher-forbids>.

## **Other objections**

Another common objection is that the Internet is already congested and has too much information, so why add to it? This is simply nonsense. The Internet does not yet contain as much information as there is in print, yet we do not worry about adding to that body of knowledge. However, everyone would welcome access to more precise and more reliable search tools to find information that they need on the Web, and it is *precisely* this problem that the OAI addresses. Searches of myOAI and OAISTER are the scholarly equivalent of Google: they search a global and growing database of information restricted to websites that advertise themselves as providing scholarly information, and which are moderated to accurately represent themselves. The value of such facilities is enhanced by increasing numbers of OAI-compliant (eprint) servers operated by universities.

Another objection that is sometimes heard is that preprint files are second-class information; the only thing that should be published is the fully refereed paper that which has been validated by experts. Such a criticism cannot be levelled against postprints or RHD theses, which eprint servers equally mount. Few editors (of which I am one) of scholarly journals would be so rash as to make this claim; the quality of the refereeing process is well-known to be patchy and is under increasing stress as more and more respected experts decline refereeing tasks. However there are two even more cogent replies.

- Eprint servers do not only provide copies of documents, they are surrounded (like e-journals and other electronic media) with other forms of validation and 'refereeing'. For example, many papers are found not through searching but by their inclusion on 'key papers' listings, links on other websites and citations. Some eprint websites also accumulate comments added to the papers; a form of democratic refereeing similar to book reviews.
- Secondly, the evidence suggests that readers do not have the same view about the uniqueness of a paper that authors sometimes do. They are often satisfied to read an earlier version of the paper if they can get it more conveniently than a later form; even more so if the author mounts a long version of a published paper. Sometimes just the abstract will satisfy them, or a text-only version without the diagrams. Enough in any case for the 200-1000 people who read the average scholarly paper to decide whether they want to study it further or order a journal published version. Both these issues are canvassed with valuable statistics about online usage in Odlyzko, A M, *Learned Publishing*, 15(1) Jan. 2002, pp7-19, also available at <http://oberon.ingentaselect.com/vl=5313651/cl=38/nw=1/fm=docpdf/rpsv/cw/alpsp/09531513/v15n1/s2/p7> (as published) and <http://www.dtc.umn.edu/~odlyzko/doc/rapid.evolution.pdf> (preprint); a 'must read' for anyone with a view on this topic, positive or negative.

It also needs to be emphasized that the eprint server outlined here has no connection with the PES (Publication Entry System) used by the Research & Development Office, though it might be possible to upload eprint data to PES records for extension with SEO, category and other information. It also has no relation to any University budget exercise in the distribution of the Research Quantum; the eprint archive places no relative values on different forms of research output.

## Recommendations

To implement a publicly accessible eprint server and get high participation as quickly as possible requires speedy implementation of some policies while others can take their time through the University system. The following recommendations provide a draft plan.

### ***General endorsement***

- R1 Academic Senate endorses the general principle of an eprint server, and requests the cooperation of the corporate sections of the University and Information Technology Resources in particular in implementing this server.

### ***Overall responsibility***

Since the implications of this scheme span the Library, research and academics generally, a distributed responsibility is desirable. The following has been discussed with the University Librarian.

- R2 Responsibility for the implementation of an eprint server and the mounting of the University's research output on it be assigned jointly to the University Librarian and the Pro Vice-Chancellor (Research).
- R3 The Librarian and the PVC(R) will be advised by a small steering committee appointed by the Academic Senate.

### ***Time-frame***

The sooner that this scheme is operational the better, as the G8 universities have a head start of about two years. Several of the existing Australian servers have been in operation for a year. R4-9 set out a desirable and achievable timeframe. Note however that the need for an eprint server is not yet built in to the University's Plans, nor the performance criteria of the individuals involved. There do arise occasions when the delays inherent in these procedures need to be bypassed, and this is one of them. This timeframe has been discussed with the University Librarian and is considered feasible.

- R4 During the remainder of 2004, the prototype server be made public and registered with OAI harvesters. The School of Computing would be the administrator and guinea-pig school in the operation of the server, but other Schools may wish to voluntarily participate (for example Information Systems).
- R5 Early in 2005, the University Library in consultation with stakeholders and the Steering Committee would implement a central server, migrate existing files, and commence to encourage voluntary participation by Schools and individuals. A target for voluntary participation by academics in uploading research documents by end 2005 is suggested as 40% of research output.
- R6 Towards the end of 2005, consideration would be given to continuing the voluntary program or making the responsibility for uploading mandatory for 2006. A target for participation by academics in uploading research documents by end 2006 is suggested as 70% of research output.
- R7 In both 2005 and 2006, policies and procedures would be put in place to meet the same percentage targets for RHD theses, though in this case the aim should be to require to submit a complete or near-complete electronic copy of the thesis to the RHD Office.
- R8 Following achievement of the 2006 target or at an earlier time considered appropriate the steering committee will be disbanded and the ongoing responsibility for the service vested in the University Library.

Appropriate policies will need to be discussed and agreed, if a high level of academic participation is to be a reality. Organization of additional workload or implementation effort will also need to be considered.

R9      Academic Senate refers this paper and the attached draft policies to the Faculties, the Board of Graduate Studies, the Tasmania University Postgraduate Association, the RHD Unit, the Web Development Office, and the Research & Development Office for discussion. Comment is to be received in time for the Senate meeting on 22 October 2004.

## Appendix 1

### Draft Research Eprint Policy

1. The University of Tasmania's policy is to maximise the visibility, usage and impact of its research output by maximising online access to the research for all would-be users and researchers worldwide.
2. It is also the University's policy to keep to the minimum the effort that each researcher has to expend in order to provide open online access to his or her research output.
3. The University has accordingly adopted the policy that all legally allowable research output is to be self-archived by researchers in a University EPrint Server.
4. This policy will be progressively implemented over the remainder of 2004 and 2005, so that by end 2005 all publicly accessible research output is uploaded to the server at the time of writing and publication. Responsibility for the implementation is assigned to the University Librarian and the Pro Vice-Chancellor (Research) jointly.
5. Publicly accessible research output includes all refereed journal articles and conference papers/short papers/poster presentations; all unrefereed conference papers/short papers/poster presentations, newspaper articles, books and research monographs, book chapters, and theses of graduating RHD candidates. Optionally, researchers may include long versions of published papers, errata, internal technical reports. Publications under a permanent or temporary embargo because of third party sponsorship are of course excluded as full-text entries, but should be included as abstracts, titles, etc.
6. Consideration should be given to altering the thesis submission rules to require the provision of an electronic version of the thesis at an appropriate time (refer to Board of Graduate Studies).
7. This archive will form a comprehensive record of the University's research publications, and the University commits to referring to it in the University's Annual Report and Research Report. [Note that the archive goes beyond the information entered into PES, which will continue for Commonwealth and budget purposes.]

### ***Advice regarding implementation***

One of the key matters for discussion is Policy 3 above. The evidence from existing archives suggests that voluntary participation yields low participation rates, and the University will fail to meet its objective. Policy 3 suggests that participation is *required* for all research output. However, this should be phased in over a year, with the Library conducting training sessions for each school, and establishing a proxy-service desirably in the school but within the Library as a backup to assist researchers who are unable to or unwilling to load their own research.

1. The University does not require the full text of ***books or research monographs*** to be uploaded. It is sufficient to archive a reference along with the usual metadata.
2. ***PhD and research Master theses*** should be archived at the point that the candidate is approved for graduation. The uploading is assigned to the RHD Unit for implementation. Thesis submission guidelines will need to be revised so that candidates provide a complete (or near complete) electronic version of the thesis in an acceptable format. Restricted access theses will not be uploaded as full text, or will be uploaded when the reason for the restriction expires.



3. ***Research papers*** submitted to journals and conferences should be uploaded as a preprint at the time of submission. Following revision, acceptance and publication, a revised record should be stored if the publisher's policies permit (see below).

This policy is compatible with publishers' copyright agreements in the following ways:

- The copyright for all unrefereed preprints resides entirely with the author(s) and the University before it is submitted for peer-reviewed publication, hence it can be self-archived irrespective of the copyright policy of the journal to which it is eventually submitted. There are no legal issues. Publishers may however exert anti-competitive coercion.
- In case you have signed a restrictive agreement in which you have voluntarily agreed to give up your IP rights and not to self-archive any preprint, you are encouraged to self-archive a version of your research findings in a long form different from the paper submission., or just a title and abstract. The IP and its disposition resides with you at this point, and copyright resides in the expression not the idea.
- The copyright for the peer-reviewed postprint will depend on the wording of the copyright agreement which you sign with the publisher, and self-archiving of the postprint will depend on this agreement.
- Many publishers allow the peer-reviewed postprint to be self-archived (eg American Psychological Association). The copyright transfer agreement will either specify this right explicitly or the author can inquire about it directly. If you are uncertain about the terms of your agreement, a table of copyright policies is available at <http://www.sherpa.ac.uk/romeo.php>. *Wherever possible, you are advised to modify your copyright agreement so that it does not disallow self-archiving.*
- In case you have signed a restrictive copyright transfer form in which you have explicitly agreed not to self-archive the peer-reviewed postprint, you are encouraged to self-archive, alongside your already-archived preprint, a 'corrigenda' file, listing the substantive changes the user would need to make in order to turn the unrefereed preprint into the refereed postprint.
- Copyright agreements may state that eprints can be archived on your personal homepage. As far as publishers are concerned, the eprint archive is a part of the University's infrastructure for your personal homepage.
- Some journals still maintain submission policies which state that a preprint will not be considered for publication if it has been previously 'publicised' by making it accessible online. Unlike copyright transfer agreements, such policies are not a matter of law but simple coercion by the publisher. If you have concerns about submitting an archived paper to a journal which maintains such a restrictive submission policy, please discuss the matter with the University's IP Adviser.

## Appendix 2

### Draft Research Higher Degree Thesis Policy

This draft policy does not pretend to be explicit, nor final. Rather it sets out some desirable principles for discussion from which a policy can be formed. The intention is to enhance the visibility of theses, which is a win-win situation for both the graduate and the University. Text in italics at the end of a point is an explanation.

1. During 2005, RHD candidates will be encouraged to submit their RHD thesis in an appropriate electronic form as well as hard copy. *[There must be few if any candidates who still submit typewritten theses, so the only issues are whether all the components are assembled into a single e-form or whether the paper thesis is assembled from printouts of various files, photographs, and other records. A trial year will provide time for experience, and the major benefits of eprint archiving should be an incentive for some to participate.]*
2. From 2006, submission of an electronic form of the thesis will be mandatory. *[If the e-form is mandatory, its production becomes part of the thesis writing and production, and simplified because it is no longer an "add-on". Photographs and other information will also be e-assembled rather than paper cut and pasted (or inserted as pages). Significant advisories will need to be developed, and maybe there should remain a provision for "near-complete" e-forms.]*
3. "Near complete" e-forms should be acceptable. As the name suggest a near complete e-form contains the vast bulk of the thesis, certainly all the text, but may be missing some parts which are considered too hard to integrate, such as a photocopy of a published article, a complex set of high-resolution photographs, etc. *[Such e-forms are likely to be almost as valuable to the reader as the full thesis, which can then be retrieved if the missing pieces are desired. The eprint record would identify the document as "incomplete" and give details.]*
4. At some time in the future (possibly 2006) submission of a hard copy for the University Library will be no longer required. *[The only justification of shelving hard copies of theses once the e-form is mandated to be complete, is for the archival value of paper records. They are hardly ever accessed at present and will be even less so once e-forms are available.]*
5. The Board of Graduate Studies and the RHD Unit will arrange for appropriate training and assistance to candidates in developing their publishing skills to make this possible. *[It should be seen as an essential part of research training to be able to produce a research monograph in e-form.]*
6. Consideration should be given to allowing candidates to optionally submit their theses only in electronic form from 2006. *[This has workload benefit for the candidates in producing only one form, and mirrors the real world where few researchers assemble paper copy any more. In addition there are major benefits for the candidate in being able to use colour diagrams, graphics and photographs, and even to include animations and videos. The option might be used sparingly at first in frontier technologies, and it is too early to make e-submission the norm though some universities are moving this way.]*